Claims

1. A method for the co-articulation-specific concatenation of audio segments, in order to generate synthesised acoustical data which reproduces a sequence of concatenated sounds/ phones, comprising the following steps:
- selection of at least two audio segments which contain bands, each of which reproducing a portion of a sound/phone or a portion of a sound/phone sequence,
characterised by the steps of:
- establishing a band to be used of an earlier audio segment;
- establishing a band to be used of a later audio segment, which begins immediately before the band to be used of the later audio segment and ends with the co-articulation band of the later audio segment which follows the initially used solo articulation band;
- with the duration and position of the bands to be used being determined as a function of the earlier and later audio segments; and
- concatenating the established band of the earlier audio segment with the established band of the later audio segment, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in its established band.

2. The method according to Claim 1, characterised in that
- the instance of concatenation is set in a band which lies in the vicinity of the boundaries of the initially to be used solo articulation band of the later audio segment, if the band of same to be used reproduces a static sound/phone at the beginning; and
- a downstream portion of the band to be used of the earlier audio segment and an upstream portion of the band to be used of the later audio segment are processed by means of suitable

transfer functions and added in an overlapping manner (cross fade), with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

3. The method according to Claim 1, characterised in that
- the instance of concatenation is set in a band which lies immediately before the band to be used of the later audio segment, if the used band of same reproduces a dynamic sound/phone at the beginning; and
- a downstream portion of the band to be used of the earlier audio segment and an upstream portion of the band to be used of the later audio segment are processed by means of suitable transfer functions and joined in a non-overlapping manner (hard fade), with the transfer functions being determined depending on the acoustical data to be synthesised.

4. The method according to one of Claims 1 to 3, characterised in that for a sound/phone or a portion of the sequence of concatenated sounds/phones at the start of the concatenated sound/phone sequence a band of an audio segment is selected so that the start of the band reproduces the properties of the start of the concatenated sound/phone sequence.

5. The method according to one of Claims 1 to 4, characterised in that for a sound/phone or a portion of the sequence of concatenated sounds/phones at the end of the concatenated sound/phone sequence a band of an audio segment is selected so that the end of the band reproduces the properties of the end of the concatenated sound/phone sequence.

6. The method according to one of Claims 1 to 5, characterised in that the voice data to the synthesised is combined in groups, each of which being described by an individual audio segment.

7. The method according to one of Claims 1 to 6, characterised in that an audio segment is selected for the later audio segment band, which reproduces the highest number of successive portions of the sounds/phones of the sound/phone sequence, in order to use the smallest number of audio segment bands in the generation of the synthesised acoustical data.

8. The method according to one of Claims 1 to 7, characterised in that a processing of the used bands of individual audio segments is carried out by means of suitable functions depending on properties of the concatenated sound/phone sequence, with these properties involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

9. The method according to one of Claims 1 to 8, characterised in that a processing of the used bands of individual audio segments is carried out by means of suitable functions in a band, in which the instance of concatenation lies. This can include i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

10. The method according to one of Claims 1 to 9, characterised in that the instance of concatenation is set in places of the bands to be used of the earlier and/or later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero point, amplitude value, gradient, derivative of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

11. The method according to one of Claims 1 to 10, characterised in that

- the selection of the used bands of individual audio segments, their processing, their variation, as well as their concatenation are additionally carried out with the application of heuristic knowledge which is obtained by an additionally carried out heuristic method.

12. The method according to one of Claims 1 to 11, characterised in that
- the acoustical data to be synthesised is voice data, and the sounds are phones;
- the static phones include vowels, diphtongs, liquids, vibrants, fricatives and nasals; and
- the dynamic phones include plosives, affricates, glottal stops, and click sounds.

13. The method according to one of Claims 1 to 12, characterised in that
- a conversion of the synthesised acoustical data to acoustical signals and/or voice signals is carried out.

14. A device for the co-articulation-specific concatenation of audio segments, in order to generate synthesised acoustical data which reproduces a sequence of phones, comprising:
- a database in which audio segments are stored, each of which reproducing portion of a phone or portions of a sequence of (concatenated) phones;
- and/or any upstream synthesis means (not part of this invention) which supplies audio segments;
- a means for the selection of at least two audio segments from the database and/or the upstream synthesis means; and
- a means for the concatenation of audio segments, characterised in that the concatenation means is suited for
- defining a band to be used of an earlier audio segment;
- defining a portion to be used of a later audio segment in a band which starts with the later audio segment and ends after

a co-articulation band of the later audio segment, which follows after the initially used solo articulation band;
- determining the duration and position of the used bands depending on the earlier and later audio segments; and
- concatenating the used band of the earlier audio segment with the used band of the later audio segment by defining the instance of concatenation as a function of properties of the used band of the later audio segment in a band which starts immediately before the used band of the later audio segment and ends with the co-articulation band which follows after the initially used solo articulation band after of the later audio segment.

15. The device according to Claim 14, characterised in that the concatenation means comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment, whose used band reproduces a static phone at the beginning in the vicinity of the boundaries of the initially occurring solo articulation band of the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment by suitable transfer functions; and
- means for the overlapping addition of the two bands in an overlapping portion (cross fade), which depends on the audio segments to be concatenated, with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the acoustical data to be synthesised.

16. The device according to Claim 14, characterised in that the concatenation means comprises:
- means for the concatenation of the used band of the earlier audio segment with the used band of the later audio segment,

whose used band reproduces a dynamic phone at the beginning, immediately before the used band of the later audio segment;
- means for processing a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment by suitable transfer functions, with the transfer functions being determined depending on the acoustical data to be synthesised; and
- means for the non-overlapping joining of the two audio segments.

17. The device according to one of Claims 14 to 16, characterised in that the database includes audio segments or the upstream synthesis means supplies audio segments which comprise bands which at the start reproduce a phone or a portion of the concatenated phone sequence at the start of the concatenated phone sequence.

18. The device according to one of Claims 14 to 17, characterised in that the database includes audio segments or the upstream synthesis means supplies audio segments which comprise bands, whose ends reproduce a phone or a portion of the concatenated phone sequence at the end of the concatenated phone sequence.

19. The device according to one of Claims 14 to 18, characterised in that the database includes a group of audio segments or the upstream synthesis means supplies audio segments which comprise bands, whose starts each reproduce only a static phone.

20. The device according to one of Claims 14 to 19, characterised in that the concatenation means comprises:
- means for the generation of further audio segments by concatenation of audio segments, with the starts of the bands each reproducing a static phone, each with a band of a later

audio segment whose used band reproduces a dynamic phone at the start, and
- a means which supplies the further audio segments to the database or the selection means.

21. The device according to one of Claims 14 to 20, characterised in that, in the selection of the audio segment bands from the database or the upstream synthesis means, the selection means is suited to select the audio segments which reproduce the greatest number of successive portions of concatenated phones of the concatenated phone sequence.

22. The device according to one of Claims 14 to 21, characterised in that the concatenation means comprises means for processing the used bands of individual audio segments with the aid of suitable functions, depending on properties of the concatenated phone sequence. Among others, this can be a modification of the frequency, the duration, the amplitude, or the spectrum.

23. The device according to one of Claims 14 to 22, characterised in that
- the concatenation means comprises means for processing the used bands of individual audio segments with the aid of suitable functions in a band including the instance of concatenation, with this function involving i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

24. The device according to one of Claims 14 to 23, characterised in that
- the concatenation means comprises means for the selection of the instance of concatenation in a place in the used bands of the earlier and/or the later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.:

zero point, amplitude value, gradient, derivatives of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

25. The device according to one of Claims 14 to 24, characterised in that
- the selection means comprises means for the implementation of heuristic knowledge which relates to the selection of the used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

26. The device according to one of Claims 14 to 25, characterised in that
- the database includes audio segments or the upstream synthesis means supplies audio segments which include bands, each of which reproducing at least a portion of a sound or phone, respectively, a sound or phone, respectively, portions of phone sequences or polyphones, respectively, or sound/phone sequences or polyphones, respectively, with a static sound corresponding to a static phone and comprising vowels, diphtongs, liquids, vibrants, fricatives, and nasals; and
a dynamic sound corresponding to a dynamic phone and comprising plosives, affricates, glottal stops, and klick speech; and
- the concatenation means is suitable to generate synthesised voice data by means of the concatenation of audio segments.

27. The device according to one of Claims 14 to 26, characterised in that
- means are provided for the conversion of the synthesised acoustical data to acoustical signals and/or voice signals.

28. Synthesised voice signals which consist of a sequence of sounds or phones, respectively, with the voice signals being generated in that:

- at least two audio segments are selected which reproduce the sounds or phones, respectively; and
- the audio segments are linked by a co-articulation-specific concatenation, with
- a band to be used of an earlier audio segment being established;
- a band to be used of a later audio segment being established which starts immediately before the band to be used of the later audio segment and ends with the co-articulation band of the later audio segment, following the initially used solo articulation band;
- with the duration and position of the bands to be used being determined depending on the audio segments; and
- the used bands of the audio segments being concatenated in a co-articulation-specific manner, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in its established band.

29. The synthesised voice signals according to Claim 28, characterised in that the voice signals are generated in that
- the audio segments are concatenated in an instance which lies in the vicinity of the boundaries of the initially occurring solo articulation band of the used band of the later audio segment, if the start of this band reproduces a static sound or a static phone, respectively, with the static phone being a vowel, a diphtong, a liquid, a fricative, a vibrant, or a nasal; and
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by means of suitable transfer function and both bands are added in an overlapping manner (cross fade), with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

30. The synthesised voice signals according to Claim 28, characterised in that the voice signals are generated in that
- the audio segments are concatenated in an instance which lies immediately before the used band of the later audio segment, if the start of this band reproduces a dynamic sound or phone, respectively, with the dynamic phone being a plosive, an affricate, a glottal stop, or klick speech; and
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by means of suitable transfer functions and both bands are joined in a non-overlapping manner (hard fade), with the transfer functions being determined depending on the audio segments to be concatenated.

31. The synthesised voice signals according to one of Claims 28 to 30, characterised in that
- the first sound or the first phone, respectively, or a portion of the first phone sequence or of the first polyphone, respectively, in the sequence is generated by an audio segment, whose used band at the start reproduces the properties of the start of the sequence.

32. The synthesised voice signals according to one of Claims 28 to 30, characterised in that
- the last sound or the last phone, respectively, or a portion of the last phone sequence or of the last polyphone, respectively, in the sequence is generated by an audio segment, whose used band at the end reproduces the properties of the end of the sequence.

33. The synthesised voice signals according to one of Claims 28 to 32, characterised in that
- the voice signals are generated in that later bands of audio segments, beginning with the reproduction of a dynamic sound or phone, respectively, are concatenated with earlier bands of

audio segments, beginning with the reproduction of a static sound or phone, respectively.

34. The synthesised voice signals according to one of Claims 28 to 33, characterised in that
- such audio segments are selected which reproduce the highest number of portions of sounds or phones, respectively, of the sequence, in order to use the smallest number of audio segment bands in the generation of the voice signals.

35. The synthesised voice signals according to one of Claims 28 to 34, characterised in that
- the voice signals are generated by the concatenation of the used bands of audio segments which are processed with the aid of suitable functions depending on properties of the sound sequence or phone sequence, respectively. This can include i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

36. The synthesised voice signals according to one of Claims 28 to 35, characterised in that
- the voice signals are generated by the concatenation of the used bands of audio segments which are processed with the aid of suitable functions depending on properties of the sound sequence or phone sequence, respectively, in band in which the instance of concatenation lies, with these properties including i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

37. The synthesised voice signals according to one of Claims 28 to 36, characterised in that the instance of concatenation lies at a place in the used bands of the earlier and/or the later audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero point, amplitude value,

gradient, derivative of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

38. The synthesised voice signals according to one of Claims 28 to 37, characterised in that the voice signals are suited for a conversion to acoustical signals.

39. A data carrier which includes a computer program for the co-articulation-specific concatenation of audio segments in order to generate synthesised acoustical data which reproduces a sequence of concatenated phones, comprising the following steps:
- selection of at least two audio segments which contain bands, each of which reproducing a portion of a sound/phone or a portion of a sound/phone sequence,
characterised by the steps of:
- establishing a band to be used of an earlier audio segment;
- establishing a band to be used of a later audio segment, which begins immediately before the band to be used of a later audio segment and ends with the co-articulation band of the later audio segment which follows the initially used solo articulation band;
- with the duration and position of the bands to be used being determined as a function of the earlier and later audio segments; and
- concatenating the established band of the earlier audio segment with the established band of the later audio segment, in that the instance of concatenation, as a function of properties of the used band of the later audio segment, is set in its established band.

40. The data carrier according to Claim 39, characterised in that the computer program selects the instance of the conca-

tenation of the used band of the second audio segment with the used band of the first audio segment in such a manner that
- the instance of concatenation is set in a band which lies in the vicinity of the boundaries of the initially used solo articulation band of the later audio segment, if its used band reproduces a static phone at the start;
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by suitable transfer functions and added in an overlapping manner (cross fade), with the transfer functions and the length of an overlapping portion of the two bands being determined depending on the audio segments to be concatenated.

41. The data carrier according to Claim 39, characterised in that the computer program selects the instance of the concatenation of the used band of the second audio segment with the used band of the first audio segment in such a manner that
- the instance of concatenation is set in a band which lies immediately before the used band of the later audio segment, if the used band of same reproduces a dynamic sound/phone at the beginning; and
- a downstream portion of the used band of the earlier audio segment and an upstream portion of the used band of the later audio segment are processed by means of suitable transfer functions and joined in a non-overlapping manner (hard fade), with the transfer functions being determined depending on the audio segments to be concatenated.

42. The data carrier according to one of Claims 39 to 41, characterised in that the computer program selects a band of an audio segment for a phone or a portion of the sequence of concatenated phones at the start of the concatenated phone sequence, the start of which reproduces the properties of the start of the concatenated sequence of phones.

43. The data carrier according to one of Claims 39 to 42, characterised in that the computer program selects a band of an audio segment for a phone or a portion of the sequence of concatenated phones at the end of the concatenated phone sequence, the end of which reproduces the properties of the end of the concatenated sequence of phones.

44. The data carrier according to one of Claims 39 to 43, characterised in that the computer program carries out a processing of the used bands of individual audio segments with the aid of suitable functions depending on properties of the phone sequence. This can include i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

45. The data carrier according to one of Claims 39 to 44, characterised in that the computer program selects an audio segment band for the later audio segment band which reproduces the highest number of successive portions of the concatenated phones in the phone sequence, in order to use the smallest number of audio segment bands in the generation of the synthesised acoustical data.

46. The data carrier according to one of Claims 39 to 45, characterised in that the computer program carries out a processing of the used bands of individual audio segments with the aid of suitable functions in a band in which the instance of concatenation lies. This can include i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

47. The data carrier according to one of Claims 39 to 46, characterised in that the computer program establishes the instance of concatenation in a place of the used bands of the first and/or the second audio segment, in which the two used bands are in agreement with respect to one or several suitable properties, with these properties including i.a.: zero point,

amplitude value, gradient, derivative of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

48. The data carrier according to one of Claims 39 to 47, characterised in that the computer program carries out an implementation of heuristic knowledge which relates to the selection of the used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

49. The data carrier according to one of Claims 39 to 58, characterised in that the computer program is suited for the generation of synthesised voice data, with the sounds being phones, the static phones comprising vowels, diphtongs, liquids, vibrants, fricatives, and nasals; and the dynamic phones comprising plosives, affricates, glottal stops, and klick speech.

50. The data carrier according to one of Claims 39 to 49, characterised in that the computer program converts the synthesised acoustical data to acoustical convertible data and/or voice signals.

51. An acoustical, optical, magnetic, or electrical data storage which contains audio segments in order to generate synthesised acoustical data by means of a concatenation of used bands of the audio segments, utilising the methods according to Claim 1, or the device according to Claim 14, or the data carrier according to Claim 39.

52. The data storage according to Claim 51, characterised in that a group of the audio segments reproduces sounds or

phones, respectively, or portions of sounds or phones, respectively.

53. The data storage according to Claim 51 or 52, characterised in that a group of the audio segments reproduces phone sequences or portions of phone sequences or polyphones, respectively, or portions of polyphones.

54. The data storage according to one of Claims 51 to 53, characterised in that a group of audio segments is provided whose used bands start with a static sound or phone, respectively, with the static phones comprising vowels, diphtongs, liquids, fricatives, vibrants, and nasals.

55. The data storage according to one of Claims 51 to 54, characterised in that audio segments are provided which are suitable for the conversion to acoustical signals.

56. The data storage according to one of Claims 51 to 55, which additionally contains information in order to carry out a processing of the used bands of individual audio segments with the aid of suitable functions depending on properties of the acoustical data to be synthesised. This can be i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

57. The data storage according to one of Claims 51 to 56, which additionally contains information relating to a processing of the used bands of individual audio segments with the aid of suitable functions in a band in which the instance of concatenation lies. This can be i.a. a modification of the frequency, the duration, the amplitude, or the spectrum.

58. The data storage according to one of Claims 51 to 57, which additionally provides linked audio segments, whose in-

stance of concatenation lies at a place of the used bands of the earlier and/or later audio segment, where both used bands are in agreement with respect to one or several suitable properties. These properties can be i.a.: zero point, amplitude value, gradient, derivative of any degree, spectrum, tone level, amplitude value in a frequency band, volume, style of speech, emotion of speech, or other properties covered in the phone classification scheme.

59. The data storage according to one of Claims 51 to 58, which additionally contains information in the form of heuristic knowledge, which relates to the selection of the used bands of the individual audio segments, their processing, their variation, as well as their concatenation.

60. Sound carrier which contains data which at least partially is synthesised acoustical data which were generated
- by means of a method according to one of Claims 1 to 13, or
- by means of a device according to one of Claims 14 to 27, or
- by utilising a data carrier according to one of Claims 39 to 50, or
- by utilising a data storage according to one of Claims 51 to 59, or
- which are the voice signals according to one of Claims 28 to 38.

61. The sound carrier according to Claim 60, characterised in that the synthesised acoustical data is synthesised voice data.